Sensitivity and specificity of SkinVision are likely to have been overestimated

Deeks, J J; Dinnes, J; Williams, H C

DOI:
10.1111/jdv.16382

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Deeks, JJ, Dinnes, J & Williams, HC 2020, 'Sensitivity and specificity of SkinVision are likely to have been overestimated', *Journal of the European Academy of Dermatology and Venereology*, vol. 34, no. 10, pp. e582-e583. https://doi.org/10.1111/jdv.16382

LETTER TO THE EDITOR

Sensitivity and specificity of SkinVision are likely to have been overestimated

Editor

Udrea *et al.*[1] reported 95.1% sensitivity and 78.3% specificity using the SkinVision app for risk assessment of skin lesions. These are likely to be overestimated due to the nature of the study design and flaws in sampling.

The study design, where inclusion of those with and without disease is undertaken separately, is known as a diagnostic case–control study or a two-gate design.[2] These studies have been shown to overestimate both sensitivity and specificity across many clinical applications compared to the ideal design, where all using the test as intended are recruited to a study prior to ascertaining disease status.[3]

The risk of bias in this study is further increased by the method of lesion selection. The app is intended for use by laypersons to self-assess skin lesions; thus, its accuracy should be assessed in this context. Three previous studies of SkinVision have been entirely based on clinically selected cohorts.[4–6] Udrea *et al.* did use some app users' data for evaluation; however, these data are mixed with clinical cohorts, and the app users have been selected in a way that is almost certain to have introduced bias.

The sensitivity of the app was evaluated in two cohorts of clinically investigated patients combined with a cohort of app users. The clinical cohorts are from previous studies[5,6] including 48 malignant melanomas (40 from Munich,[5] 8 from Eindhoven[6]) with 147 other malignant and premalignant lesions (all from Eindhoven: 107 basal and 8 squamous cell carcinomas; 18 actinic keratosis; 14 Bowen's Disease). Although inclusion of secondary care cohorts ensures histological verification, these patients have gone through several selection steps (choosing to present to primary care, referral from primary care, excision decision in secondary care) and are thus unlikely to be representative of typical app users, and possibly reflect more clearly defined cases than the early skin cancer that the app aims to detect.

However, it is the selection of data collected from app users which is the more serious flaw. All app users had used a previous SkinVision app, with the image checked by a dermatologist – a senior dermatologist for high-risk images and a junior dermatologist for low-risk images. Users with high-risk results were invited to provide follow-up data, resulting in the inclusion of 90 high-risk-rated melanomas. Malignant lesions previously missed by the app (estimated as 20% in the Eindhoven cohort[6]) and the check by junior dermatologists could not have been included and are likely to be the more difficult to diagnose cases. The proportion of those invited to follow-up who provided histology results (338 of 48 547 – 0.7%) is very small, with substantial further potential for selection bias. Verification of the 6000 benign lesions was based on the junior dermatologist's assessment of a single image taken by the app with no histology or follow-up data, which is suboptimal. Recruitment of benign lesions seems to have been restricted to those judged as low or medium risk by the previous SkinVision app and clinical assessment, and the false-positive cases from that app (22% of benign lesions in the Eindhoven data[6]) will have been automatically excluded, and are likely to have been more difficult to diagnose. These flaws make overestimation of both sensitivity and specificity highly likely.

Whilst skin cancer diagnosis apps might make valuable contributions, screening applications for public use can easily do harm through missing skin cancer cases and from over-investigation of false positives. It is essential that appropriate studies are done before recommendations are made to widely disseminate technologies, as is happening with SkinVision.[7] We have recently reviewed all evidence of the accuracy of skin diagnosis apps which use inbuilt algorithms and failed to find any examples of well-done studies.[8] Our review contains constructive guidance for investigators to minimize bias and increase external validity to app users – we hope that future studies can meet these criteria.

Funding source

None.

Conflict of Interest

All authors indicate that they have no conflicts of interest to declare.

J.J. Deeks,[1,2,*] J. Dinnes,[1,2] H.C. Williams[3]

[1]Test Evaluation Research Group, Institute of Applied Health Research, University of Birmingham, Birmingham, UK, [2]NIHR Birmingham Biomedical Research Centre, University Hospitals Birmingham NHS Foundation Trust and University of Birmingham, Birmingham, UK, [3]Centre of Evidence-Based Dermatology, Nottingham University Hospitals NHS Trust, Nottingham, UK

*Correspondence: J.J. Deeks. E-mail: j.deeks@bham.ac.uk

References

1 Udrea A, Mitra GD, Costea D *et al.* Accuracy of a smartphone application for triage of skin lesions based on machine learning algorithms. *J Eur Acad Dermatol Venereol* 2019; **34**: 648–655.

2 Rutjes AW, Reitsma JB, Vandenbroucke JP, Glas AS, Bossuyt PM. Case-control and two-gate designs in diagnostic accuracy studies. *Clin Chem* 2005; **51**: 1335–1341.
3 Lijmer JG, Mol BW, Heisterkamp S *et al*. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999; **282**: 1061–1066.
4 Chadwick X, Loescher LJ, Janda M, Soyer HP.Mobile medical applications for melanoma risk assessment: false assurance or valuable tool? 47th Hawaii International Conference on System Sciences; 6–9 January, 2014; 2675–2684.
5 Maier T, Kulichova D, Schotten K *et al*. Accuracy of a smartphone application using fractal image analysis of pigmented moles compared to clinical diagnosis and histological result. *J Eur Acad Dermatol Venereol* 2015; **29**: 663–667.
6 Thissen M, Udrea A, Hacking M, von Braunmuehl T, Ruzicka T. mHealth app for risk assessment of pigmented and non-pigmented skin lesions-a study on sensitivity and specificity in detecting malignancy. *Telemed J E Health* 2017; **23**: 948–954.
7 SkinVision. [WWW document]. URL https://www.skinvision.com/news/skinvision-to-join-nhs-innovation-accelerator (last accessed: 26 May 2020).
8 Freeman K, Dinnes J, Chuchu N *et al*. Algorithm based smartphone apps to assess risk of skin cancer in adults: systematic review of diagnostic accuracy studies. *BMJ* 2020; **368**: m127.

DOI: 10.1111/jdv.16382